Social Network Analysis and Mining (SNAM) manuscript No.
(will be inserted by the editor)

ArAutoSenti: Automatic annotation and new tendencies for sentiment classification of Arabic messages

**Imane Guellil · Faical Azouaou ·
and Francisco Chiclana**

the date of receipt and acceptance should be inserted later

Abstract A corpus-based sentiment analysis approach for messages written in Arabic and its dialects is presented and implemented. The originality of this approach resides in the automation construction of the annotated sentiment corpus, which relies mainly on a sentiment lexicon that is also constructed automatically. For the classification step, shallow and deep classifiers are used with features being extracted applying word embedding models. For the validation of the constructed corpus, we proceed with a manual reviewing and it was found that 85.17% were correctly annotated. This approach is applied on the under-resourced Algerian dialect and the approach is tested on two external test corpora presented in the literature. The obtained results are very encouraging with an F1-score that is up to 88% (on the first test corpus) and up to 81% (on the second test corpus). These results respectively represent a 20% and a 6% improvement, respectively, when compared with existing work in the research literature.

Imane Guellil
Aston University and Folding Space
Birmingham, United Kingdom
i.guellil@aston.ac.uk
Laboratoire des Méthodes de Conception des Systèmes. Ecole nationale Supérieure d'Informatique, BP 68M, 16309, Oued-Smar, Alger, Algérie. http://www.esi.dz

Faical Azouaou
Laboratoire des Méthodes de Conception des Systèmes. Ecole nationale Supérieure d'Informatique, BP 68M, 16309, Oued-Smar, Alger, Algérie. http://www.esi.dz

Francisco Chiclana
Institute of Artificial Intelligence (IAI), Faculty of Computing, Engineering and Media, De Montfort University, Leicester, UK (chiclana@dmu.ac.uk) and Andalusian Research Institute on Data Science and Computational Intelligence, University of Granada, Spain (inv.chiclana@ugr.es).

Keywords Arabic sentiment analysis, Arabic and its dialects, Automatic resources construction, shallow/deep classification, word embedding, document embedding

1 Introduction

Opinions on a product, a company or a political personality are important for business managers and company directors. The emergence of the internet and social media has made available, and will continue to do so in the future, large amounts of data containing significant numbers of opinions, sentiments and emotions, thus engendering interest on their analysis. Sentiment analysis (SA) research has really paid off for languages such as English, French or Chinese, with frequent and numerous studies and work being published. For other languages, such as Arabic and its dialects, research works have just started to give usable results. The reason for the low amount of work focusing on the Arabic language is twofold: (1) the morphological richness of this language and its dialects makes its analysis very complex and, therefore, challenging; and (2) the lack of resources dedicated to this language and, in particular, to its dialects.

There are two main approaches proposed for Arabic SA: 1) by constructing or using a sentiment lexicon (lexicon-based approach) and 2) by constructing or using an annotated corpus (corpus-based approach). Different techniques are used for lexicon and corpus constructions, respectively. Some of them are manual while others are automatic. Almost all the lexicons have been constructed automatically based mainly on Google translate or bilingual dictionaries [91, 110, 92, 1, 5]. However, Google translate works only for Modern Standard Arabic (MSA) whilst bilingual dictionaries cover only the most studied Arabic dialects such as Egyptian and Levantine. Dialects such as the Maghrebi ones (Algerian, Tunisia and Moroccan) cannot be handled with Google translate, with only few resources available for them [67, 126]. Regarding corpus construction, almost all the available corpora have been constructed manually [106, 3, 85, 95, 2, 94] and the annotation is done by native annotators, which is time and effort consuming. Additionally, this approach produces corpus with only few thousands messages, i.e. limited corpus.

To overcome the above Arabic SA identified limitations, this paper presents an automatic lexicon-based approach construction of an annotated sentiment corpus with two principal steps:

(i) A lexicon construction that uses GlosbeAPI, which is, to the best of our knowledge, the unique API that translates from/to Arabic and all its dialects; and

(ii) The computation of a message sentiment score based on a proposed algorithm that handles the most important characteristics of Arabic and its dialect, i.e. opposition, morphological aspects and negation. When a message contains a strong emoticon such as: "<3, ☺, ☹, etc.", it is directly classified as positive or negative (depending on the valence of the used emoticon).

After constructing the corpus, Word2vec [88] and fastText [71] algorithms are used to extract features, and different shallow (Support Vector Machine (SVM) [120], Logistic Regression (LR) [32], etc.) and deep (convolutional neural network (CNN) [52], Long Short-Term Memory (LSTM) [68], etc.) classification algorithms are applied to achieve the proposed automatic classification of a set of messages[1].

The next section provides an overview of the Arabic language and its different dialects. Following this, Section 3 concerns existing literature research work on Arabic SA covering sentiment lexicon and corpora construction; new trends in Arabic SA; and the research work inspiring our approach to support our objectives and motivation. In Section 4 describes the methodology of the proposed approach of Arabic sentiment analysis using both lexicon-based and corpus-based techniques. Section 5 is devoted to its experimental evaluation. Lastly, Section 6 concludes the paper.

2 Backgroud

2.1 Sentiment Analysis: an overview

Sentiment analysis (SA), also called opinion mining, is the field of study that analyses people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions toward entities such as products, services, organisations, individuals, issues, events, topics, and their attributes. It represents an important and active research area in computer science [80]. Sentiment analysis was recently considered as a hot topic of research in social networks. [96]

Sentiment analysis can be done using two main automatic learning approaches: *supervised* and *unsupervised*. Unsupervised approaches (also called lexicon-based method) are based on sentiment lexicons like dictionaries [117,25,91, 40,55]. The lexicon-based approach relies on a sentiment lexicon containing a set of words with their valences (positive/ negative/ neutral). In addition to the valence, some lexicons also contain, the polarity of each word. The majority of authors rely on a number between 1 and 5 for the positive words. A number between (- 1) and (- 5) is used for the negative words. Then, an algorithm calculating the polarity of a given sentence is proposed. The idea of the majority of the algorithms is to summing the polarities of the different words detected in the used lexicon. The originality of each algorithm resides in its ability to rely on other features such as handling negation, intensifiers.

[1] More details related to deep learning algorithms are presented in the survey of Schmidhuber [111]

Supervised approaches (also called the corpus-based method) typically use labelled corpus to train the sentiment classifier [81,26,106,85,17]. Corpus-based approach relies on a sentiment corpus containing a set of sentences/documents (i.e. sentence level or document level) with their valences (positives/ negatives or neutral). In contrast to the lexicon-based approach, almost all the constructed corpora contain only the valence (i.e. the polarity is not used). This approach is based on a set of classification algorithms (such as both shallow and deep algorithms). The first idea of these algorithms is to train a model based on the annotated used corpus. Afterwards, the generated model is used to classify the utterance sentence into positive/ negative or neutral [60,62].

2.2 Arabic and its dialects: an overview

Arabic is the official language of 27 countries; it is spoken by more than 400 million people; and it is recognised as the 4^{th} most used language in the Internet [27,58]. The research work in the literature distinguishes three main varieties of Arabic [64,49,67]: 1) Classical Arabic (CA) used in literary texts (such as Quran – Muslims's book); 2) Modern Standard Arabic (MSA) used for writing as well as formal conversations; and 3) Dialectal Arabic (DA) used in daily life communication and conversational level [27].

Arabic Dialects are another form of Arabic used in daily life communication that is referred to by the term 'dārija', which means 'current language'. Almost all Arab countries have their own dialects, which can differ within the same country [107]. The difference between dialects are due to the history of each country and their geographical locations. For example, the Algerian dialect has been influenced over the centuries by languages such as Amazigh, Turkish, Italian, Spanish and French. For example, the following words قرجومه *Qurjuwmah* 'gorge', زبله *Zablah* 'fault', سبردينه *Spardinyah* 'Espadrille' and تيليفون *Tiyliyfuwn* 'Telephone' are borrowed from the Berber, Turkish, Italian, Spanish and French languages, respectively [108]. Arabic dialects are mostly divided into six main groups: (1) Egyptian (EGY), which is the most widely understood dialect, due to the spreading of the Egyptian television and movie industry; (2) Levantine (LEV), which represents a set of dialects that differ in pronunciation and intonation, but that are largely equivalent in written forms and closely related to the Aramaic language; (3) Gulf (GLF), which is the closest regional dialect to MSA; (4) Iraqi (IRQ), which is considered to be a Gulf dialect with its own distinctive features in terms of prepositions, verb conjugation and pronunciation; (5) Maghrebi (MAGH), which is influenced by the French and Berber languages; and (6) Others remaining dialects [64,109,126, 59].

3 Arabic sentiment analysis: Related work

This section concerns existing research work on Arabic SA with a focus on sentiment lexicon and corpora construction; new trends in Arabic SA; and the research works inspiring our proposed approach.

3.1 Work on sentiment lexicon and corpora construction

Three research trends have emerged for the Arabic sentiment lexicon:

1. Manual lexicon construction as reported in [1,83], with the first work describing the process of the manual creation of the lexicon SIFAAT while the second work focused on the Algerian dialect and the construction of a lexicon by manual translation of an existing MSA and Egyptian lexicon.
2. Automatic lexicon construction, used for most of the Arabic sentiment lexicons presented in the literature, following one of the following three main methodologies:
   (i) Construction based on automatic translation of an existing English sentiment lexicon, such as Bing Liu's lexicon [36], SentiWordnet [48] or SentiStrenght [117], using Google translate [91,110,92,1,5] or an Arabic/English dictionary [7].
   (ii) Construction based on resources linking [22,55,47,15] such as English/Arabic resources such as Sentiwordnet, Arabic WordNet [51] and Arabic Morphological Analyzer [56,28]. Manual annotation provides a high precision but it lacks coverage while the automatic construction from existing resources offers high coverage but a lower precision [54].
   (iii) construction based on both translation and resources linking [82,7] using a reduce seed set of English sentiment words that are translated into Arabic and expanded using Arabic Wordnet or Arabic synonym dictionaries.
3. Semi-automatic lexicon construction, i.e. automatic construction of the lexicon followed by its manual review [40,4], is the least proposed methodology. For example, El-Beltagy [40] presents NileULex, an Arabic sentiment lexicon composed of 45% of words in Egyptian dialect and 55% of words in MSA. The First version of NileULex was proposed in 2013 [42]. Afterwards, new words were manually added to this lexicon [45]. Finally, the resulted lexicon was manually reviewed in order to limiting the effect of semantic ambiguity [40].

The same three research trends have also been observed for the Arabic sentiment corpus construction:

1. Manual sentiment corpus construction, which is the approach applied in the construction of almost all the Arabic sentiment corpora [106,3,85,95,2,

94, 103, 93], with the annotation being carried out, in the majority of cases, by native annotators.

2. Automatic sentiment corpus construction. Research applying an automatic construction is scarce, with the following two techniques been employed:

   (i) A rating reviews on a 1 to 5 stars scale as in [16, 46] for constructing 7 data sets; and

   (ii) A sentiment lexicon as in [58], where an Algerian sentiment lexicon was created and used for tagging a large set of Algerian messages, or as in [53], where a large sentiment corpus dedicated to MSA and Egyptian dialect was built via the manual annotation of a sentiment lexicon containing 4,404 phrases (used as keywords), which was used to propose an algorithm for automatically annotate a corpus containing more than 400,000 tweets (reduced to only 151,548 tweets where both positives and negative classes contain 75,774 tweets after preprocessing and annotation).

3. Semi-automatic sentiment corpus construction, which has been scarcely applied. The corpus AraSenTi-Tweet constructed in [11] is an example of the application of this annotation schema. It contains 17,573 Saudi tweets that were semi-automatically annotated into four classes: positive, negative, neutral and mixed. For constructing this corpus, the authors firstly relied on a sentiment lexicon containing both keywords and emoticon with their polarities (i.e. positive/negative).

The main lexicon and corpus construction approaches mentioned above are summarised in Table 1, where additional details are provided regarding their size, the literature research work using them and their link if publicly available.

3.2 New trends in Arabic SA

In a supervised approach (corpus-based approach) text is represented as a feature vector. Due to its simplicity and efficiency, a bag of words (BOW) representation model is commonly used [13]. Despite its popularity, this approach has two major drawbacks: 1) loss of word order in the sentence, and 2) semantic ignorance of words [24]. Moreover, the application of this approach may require additional treatment of data and an additional appropriate word feature extraction technique [8, 24]. Word and document embedding has emerged as an alternative representation model [8, 44, 24, 14]. Indeed, El Mahdaouy et al. [44] affirm that using document embedding improves text classification. Al-Azani and El-Alfy [8] and Altowayan et al. [14] relied on large Arabic corpora to train Word2vec models [89] in order to improve Arabic SA, while Barhoumi in [24] applied the Doc2vec model [79] for the sentiment classification of the corpus LABR [16]. More recently, the FastText algorithm has been proposed [71], which is based on either the Skip-gram or the continuous BOW (CBOW)

Resource type	Name	Size	Work using resources	link
Lexicon	SIFFAT [1]	229 452	[4,22]	NA
	Mataoui et al. lexicon [83]	3093	NA	NA
	Arabic Emoticon Lexicon + Arabic hashtag lexicon + Arabic hashtag lexicon (Dialect) + NRC Emoticon lexicon+ NRC hashtag lexicon[91,110,92]	43304 + 21964 + 20128 + 26740 + 32582	[70,77,116]	All lexicons[1]
	Abdula et al. lexicon [5]	16800	[4,22]	NA
	ArSenL [22]	33995	[21,9,50,47]	ArSenL[2]
	SLSA[47]	34821	NA	NA
	[15]	249532	NA	NA
	[82]	7400	NA	NA
	[7]	4815	NA	NA
	NileULex[40]	5953	[39,41,43]	NNileULex[3]
	SANA[4]	224564	NA	
	ArSEL[55]	32196	NA	ArSEL[4]
	SentiALG + SOCALALG[61]	3408 + 2375	[58,57]	NA
Corpus	OCA [106]	500	[82,105,18,102]	OCA[5]
	AWATIF [3]	10723	NA	NA
	TSAC [85]	17060	NA	TSAC[6]
	ASTD [95]	10000	[33,121]	ASTD + python code[7]
	DARDASHA+ TAGREED+ TAHRIR+ MONTADA[2]	2798 + 3015+ 3008 + 3097	NA	NA
	SentiAlg[58]	8000	NA	NA
	Twitter Benchmark Dataset[53]	151,548	NA	NA
	LABR[16]	63257	[10,33,14]	LABR + python code[8]
	Mourad et al. corpus [94]	2300	NA	NA
	ATT+ HTL+ MOV+ PROD+ RES1+ RES2+RES [46]	2154 + 15572 + 1524 + 4272 + 8364 + 2642+ 10970	[33]	All corpus + code[9]
	ArTwitter[6]	2000	[14,12]	ArTwitter[10]
	SANA [103]	513	NA	SANA[11]
	Arasenti-tweet[11]	17573	NA	NA
	Egyptian-tweets[93]	40000	NA	Egyptian-tweets[12]

Table 1 Lexicon and corpus Arabic sentiment resources

1 http://saifmohammad.com/WebPages/lexicons.html
2 http://www.oma-project.com/
3 https://github.com/NileTMRG/NileULex
4 http://oma-project.azurewebsites.net/
5 http://sinai.ujaen.es/oca-corpus-en/
6 https://github.com/fbougares/TSAC
7 https://github.com/mahmoudnabil/ASTD
8 https://github.com/mohamedadaly/LABR
9 https://github.com/hadyelsahar/large-arabic-sentiment-analysis-resouces
10 https://archive.ics.uci.edu/ml/datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis
11 http://rahab.e-monsite.com/medias/files/corpus.rar
12 https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/LBXV9O

architectures. Although FastText is often compared to Word2vec for classification tasks [114,112], as far as we know, it has not been used for Arabic classification.

Recently, deep learning algorithms such as a convolutional neural network (CNN), long short-term memory (LSTM) and bidirectional LSTM (Bi-LSTM) have become popular for classifying sentiments. In our research context, a scheme of Arabic sentiment classification was presented in [33] to evaluate and detect the sentiment polarity from Arabic reviews in which a CNN was trained on top of pretrained Arabic word embedding (Word2vec + CBOW + Skip-gram) for sentiment classification. In this case, the authors used the CNN architecture defined in [75], which relies on one channel that allows the adaptation of pre-trained vectors for each task, and it was applied to different corpus presented in the literature such as LABR, ASTD, etc. More recently, a model for multi-class SA using a simple Neural Network architecture of different layers was presented [19], which has the advantage of not relying on language-specific features such as anthologies, dictionaries, morphological or syntactic pre-processing. This model has been applied for English, German and Arabic languages. The Arabic language applications relied on the ASTD corpus constructed in [95]. Another new tendency in SA aims 'to retain the knowledge obtained from past learning and uses past knowledge to help future learning' [29]. In this context, Xia et al. [124] proposed a distantly supervised lifelong learning framework for large-scale social media SA, and obtained results that support the feasibility and effectiveness of this approach to deal with the challenge of continuous update of texts with dynamic topics in social media. More details about the work that have been proposed for Arabic sentiment analysis are detailed in our last survey [60].

3.3 Sentiment analysis in other languages

Almost all the resources dedicated to English sentiment analysis follow two major trends, manual and automatic construction. However the majority of the resources were constructed automatically [48,113,30,81,98,84]. Esuli et al. [48] present SentiWordNet, a lexical resource produced by using an automated classifier for associating each synset of WordNet [90] to a triplet of scores corresponding to, Positive, Negative, or Neutral. This lexicon was improved in 2010 [20]. Many recent works rely on SentiWordNet including [34,65,119, 73,37,118]. In [113], a Semantic Orientation CALculator (SO-CAL) was presented. SO-CAL includes different dictionaries of words annotated with their semantic orientation (polarity and strength) and incorporates intensifiers and negation pronouns. SO-CAL classifies text into two classes (positives, negatives). The current version of SO-CAL contains a total of 6,769 entries. Cambria et al. [30] present SenticNet, a publicly available resource for sentiment analysis exploiting AI and Semantic Web techniques. SenticNet contains more than 5,700 polarity concepts. SenticNet was used in different research works including [100,101,25].

Maas et al.[81] present a large movie review dataset extracted from IMDB[2]. This dataset was constructed automatically by using the associated binary sentiment polarity labels to each movie review. It is intended to serve as a benchmark for sentiment classification. This corpus contains 50,000 reviews split evenly into 25,000 train and 25,000 test sets. Many research works rely on IMDB dataset including [76, 115, 97, 130]. Pak et al., also propose an automatic approach to construct their annotated corpus. The authors first collect a corpus of 300,000 text posts from Twitter and split it automatically into three parts, positive, negative and objective. To collect negative and positive sentiments, they queried Twitter for two types of emoticons:(1) Happy emoticons such as: ":-)", ":)", "=)", ":D", etc. and (2) Sad emoticons such as: ":-(", ":(", "=(", ";(", etc. To collect a corpus of objective posts, they retrieved text messages from Twitter accounts of popular newspapers and magazines, such as: "New York Times", "Washington Posts", etc. For classification, the authors called different classifiers such as SVM, NB, etc. However, they attested that NB ave better results. Finally, Mcauley et al., [84] construct a dataset from Amazon. This corpus contains 35 million reviews. The data were collected by starting with a list of 75 million as in-like strings (Amazon product identifiers) obtained from the Internet Archive. This dataset is used by many research works such as [128, 87, 99]

It is worth mentioning that the automatic construction, for lexicon and corpus, has also been used for other languages such as Spanish and Romanian. To show the effectiveness of this construction technique, Banea et al. [23] carried our different experiments that relied on different English resources such as MPQA corpus [122] and OpinionFinder system [123]. They also followed a manual annotation of a sentiment corpus in Spanish and Romanian. Finally, the association of word embedding and deep learning models is also the current trends for the other languages. As examples, Chen et al., [31] incorporate user information and product information in the classification process. These authors principally rely on IMDB corpus for their training. They used word2vec for extracting vectors and LSTM model for the classification. The pair Word2vec/LSTM was also used by Dou et al., [38]. The proposed model includes into two separate parts. The first part, LSTM is applied to learn a document representation. The second part, a deep memory network containing multiple computational layers is used to predict the review rating for each document. Finally, Zhou et al., [129] present an attention-based bilingual representation learning model. The proposed model learns the distributed semantics of the documents in a source and a target languages (English was used as a source language and Chinese as a target language). In each language, LSTM network was used[3].

[2] https://www.imdb.com/

[3] A detailed survey presenting deep learning for sentiment analysis was presented by Zhang et al., [127]

3.4 The research work inspiring our proposed approach

The methodology used for constructing a sentiment lexicon in the proposed approach is inspired by previous research work on the use of Google translate to automatic translate existing English lexicons [91, 110, 92, 1] based on Arabic/English dictionaries [5]. It is noted, though, that Google translate deals with MSA only and, therefore, dialects translation is not allowed. Moreover, Arabic/English dictionaries cover MSA and the most studied dialects (Egyptian, Levantine). Hence, Glosbe API[4], which is an online API offering the translation from/to MSA and almost all its dialects, was chosen instead. Glosbe API resembles Amazon Mechanical Turk[5] but it is open source. In addition, sentiment ambiguity was addressed in our proposed approach as per the research work on semi-automatic construction in [40], which manually reviews the automatically constructed lexicon. Handling morphological aspects of Arabic dialects was approached using research work dedicated to MSA that relies on stemming tools. For example the work in [83] used the MSA designed Khoja stemmer [74] for stemming the Algerian dialect. One of the major problems is that MSA tools do not generalise well to Arabic dialects [66]. In the proposed approach herein, agglutination is treated by employing an algorithm that supports the originality of the studied dialect, and it is principally related to its prefixes, suffixes, and negative pronouns.

Proposals for constructing an annotated sentiment corpus automatically that exploit the presence of emoticons and emotions to determine the sentiment of messages can be found in [98, 69, 125] while a semi-automatic construction was proposed by Ren et al. in [104]. However these work are not dedicated to Arabic but to others languages (English, Dutch and Japanese). After a careful analysis of the Arabic text on social media, it is observed that all the emoticons are not appropriate for determining sentiment. For example, the message ‘كفى من جعل الحمقى مشاهير *:)*’, which means *Enough of making fools famous :)*, is definitely negative but contains the positive emoticon “*:)*" that represents *laugh*. Another example: احبك^احبك ^ يا اروع من رات عيني احبك*:(*, which means *You are among the most beautiful which I have seen, I love you؟ I love you؟ I love you:(*, is definitely positive but contains the negative emoticon “*:(*" that represents *sadness*. Hence general emoticons such as “*:), :(*” cannot determine, on their own, the orientation of messages. To deal with this drawback, our proposed approach considers only strong emoticons for annotation (Section 4.1.2). The work presented in [11], which relies on sentiment words for the automatic annotation of a large corpus in Saudi dialects, was also another inspirational work for us to include in our proposed methodology a sentiment algorithm for handling opposition, Arabic morphology and negation (Section 4.1.2). The work reported in [11, 8, 14, 44, 24] use Word2vec/Doc2vec for features extraction were also influential, with FastText used in our pro-

[4] https://en.glosbe.com/
[5] https://www.mturk.com/

posed approach for the purpose of comparing results. Finally, our approach was also influenced by Dahou et al. and of Attia et al. research work on CNN algorithms for sentiment classification task [33, 19]. Other algorithms such as LSTM, BiLSTM and MLP are possible and have been used in other languages. Hence, we propose to use a variety of algorithms for classification task in order to ascertain which deep learning algorithm is most suitable for Arabic sentiment classification.

4 Methodology

The main contribution of this paper is a new corpus-based SA approach for Arabic and its dialects. Figure 1 illustrates the general architecture of our proposed Arabic SA approach, with its different inputs, outputs and exchanges associated with each step are described below.

As a corpus-based approach needs an annotated corpus. Hence, one of the main aims of this contribution is to automate the corpus construction process, for which we rely on a lexicon based-approach with two key processes:

1. Arabic sentiment lexicon construction. The Arabic sentiment lexicon is constructed automatically by relying on an existing English sentiment lexicon.
2. Message score computation. The algorithm for computing message score handles the morphological aspects of Arabic and its dialects when calculating the score of messages based on the constructed lexicon.

Our proposed corpus-based approach contains three principal steps:

(1) Corpus extraction – a large Arabic corpus is firstly extracted from Facebook.

(2) Corpus annotation – each Arabic message (input) in the extracted corpus is automatically annotated as positive or negative (output) by relying mainly on the constructed Arabic sentiment lexicon.

(3) Sentiment classification – we propose an automatic classification of an Arabic message as positive or negative by using new SA trending methods like word embedding, deep learning, etc.

4.1 Lexicon-based approach

4.1.1 Arabic sentiment lexicon construction

This step receives as input an English sentiment lexicon. English is chosen because it is the most used language in SA and SA lexicon [63]. Each word in this lexicon is translated using a translation API. After that, a lexicon of

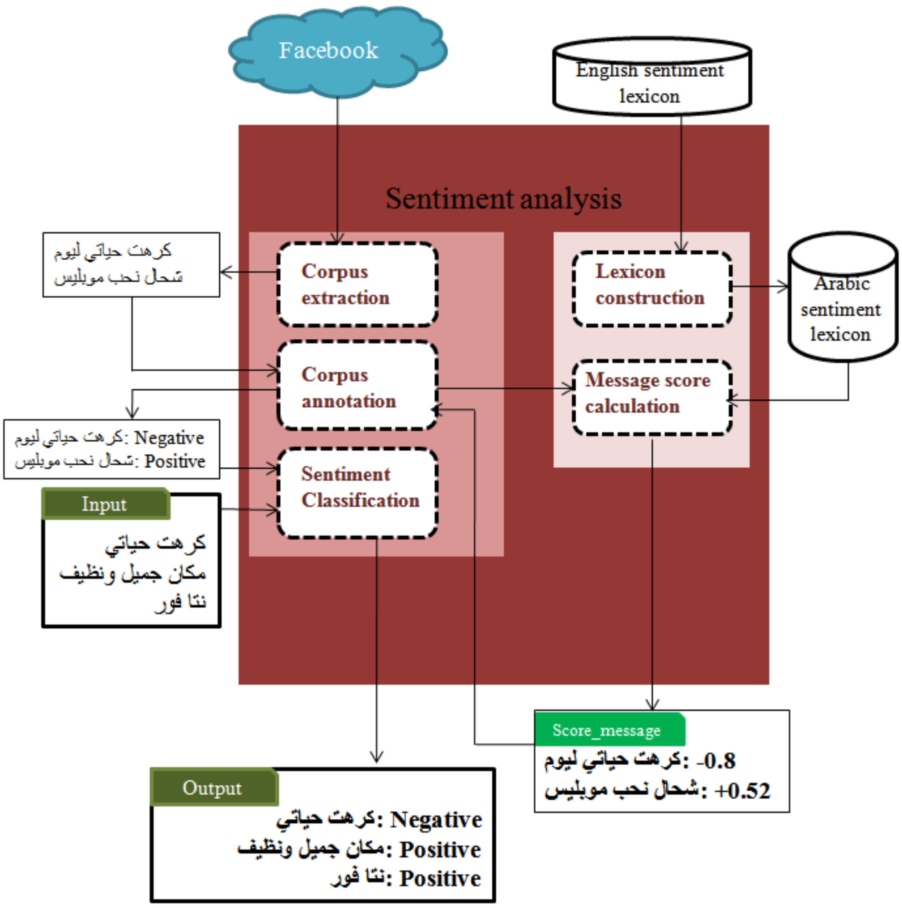

Fig. 1 The general architecture of the proposed Arabic SA approach

sentiments is constructed by extracting each term in Arabic and calculating its score. Accordingly, this step consists of two sub-steps:

4.1.1.1 Translation The Arabic lexicon is constructed by translating an existing English lexicon. Rather than SentiWordNet [20] or SentiStrength [117], SOCAL [113] was chosen because it contains a large number of terms and this study does not focus on the context of terms but only on its global valence. As already mentioned, the present work focuses on Arabic and its dialects (MSA + dialects), and therefore the Glosbe API[6] was chosen to translate English words.

4.1.1.2 Term score computation After the automatic translation, the score of an English word is associated with its translated word(s). For example, the

[6] https://glosbe.com/en/arq/excellent

score +5 of the English word 'excellent' is associated to its translations: 'باهي' (bAhy), 'لطيف' (lTyf), and 'مليح' (mlyH). As all SOCAL's terms are in the range $[-5, +5]$, the different obtained terms are also tagged from negative (labels ranging between -1 and -5) to positive (labels ranging between +1 and +5). When different English words are translated into the same Arabic word, the average sentiment score of the English words is assigned to such an Arabic word. For example, the word 'مليح' can be the translation of the English terms 'excellent' (with an associated score of +5) but also of the English term 'good' (with an associated score of +3). Hence, the Arabic term 'مليح' will have the average of sentiment scores of the corresponding translated English words.

4.1.2 Message score computation

The constructed lexicon is used to automatically provide a sentiment score for Arabic message utterances, which leads to the provision of a method to automatically construct a large sentiment training corpus. To compute the mentioned score, different steps are followed: strong emoticons and strong expression handling; opposition handling; Arabic morphology handling (agglutination); and negation handling.

4.1.2.1 Strong emoticons and strong expression handling On the one hand, only strong emoticons are used ("<3, ☺, ☹, etc."). Hence, if a message contains a strong positive emoticon then it is automatically annotated as positive; and vice versa for negative ones.On the other hand, the presence of some expressions are also crucial in determining the valence of a message. For example, Arabic people use بارك الله في (*God bless*), نموت على and (*I love a lot*) to express very positive sentiments, while they use الله يعتيهالك (*God give you bad things*) and خسارة عليك (*lost on you*) to express very negative sentiments. Hence, if a message contains a strong positive or negative expression then it is automatically annotated

Some messages contain different emoticons/ expressions. These emoticons/ expressions could also have different polarities. However, this problem is not too frequent in social media. As we focused only on a limited list of the strongest emoticons/ expressions, their use was in most of the cases uniform (i.e. having the same polarity). To validate this claim, we firstly extract from the large Arabic corpus that we collected from social media (presented in section 5.1.1, item (i)) the messages including divergence in polarity. We extracted four categories of messages: 1) messages containing both positives and negatives emoticons. 2) messages containing both positive and negative expressions. 3) messages containing both positive emoticons and negative expressions and 4) messages containing negative emoticons and positive expressions. However, the proportion of these case comparing to the totality of the corpus was very small. Only 0.012% of the total corpus contains divergent emoticons/ expressions.

The most popular case was with positive and negative emoticons representing 0.117% from the entire corpus. We also observed that in the majority of cases the first emoticon /expression found is most representative of the global polarity of the message. Hence, we annotate the messages with the polarity of the first emoticon/ expression found.

4.1.2.2 Opposition handling Opposition is generally expressed with the word *'لكن'* in MSA and with the word *'بصح'* in some dialects, which means *but, however, etc.* From the analysis of a set of messages, it can be seen that the part following the opposition word determines the valence of the message. For example, مهما صعوبة الحيات لكن **نحن مبتسيمون** (*Even with the life difficulties but* ***we are always smiling***) is considered to carry a positive sentiment despite the negative part before the opposition. The highlighted example shows that the part following the opposition is sufficient for determining the valence of messages. Thus, the aim at this stage is to determine a set of apposition words. When the system finds one of such words in a message, its sentiment score will be computed using the part of the message that follows the opposition word.

4.1.2.3 Arabic morphology handling The different morphological analysers proposed in the literature deal with a number of dialects but not the Algerian dialect. Words are typically composed of prefix(es)+stem+suffix(es), so we employ a simple rule-based light stemmer to handle Arabic prefixes and suffixes. If a word does not match any of the entries in the lexicon, all possible prefix/suffix combinations are removed to find out if the remaining possible stems would match entries in the lexicon. For example, the word نحبهالك (*I like it for you*) can be separated into ن+حب+ها+لك; the stem حب is included in the sentiment lexicon; the letter ن is included in the prefix-list; and the letters ها and لك are both included in suffix list; therefore, this splitting is accepted, and this word receive the valence and intensity of its stem حب (*love*) and the score of +1.56). Some stems that end with ى (Y) when they are in isolation, such as بكى (*cried*), when suffixes are attached to it are transformed into ي (y) , such as بكيت (*I cried*). Thus, in the context of this work, ى was normalised to ي. However, negation prefixes and suffixes are handled separately.

The prefixes / suffixes that we used are: و (w), ا (A), ي (y), ت (t), ن (n), ب (b), ال (Al), and ل (l). As for the suffixes, we used: ي (y), ت (t), و (w), ا (A), ة (p), ين (yn), ا (A), ه (h), هم (hm), كم (km), نا (nA), ها (hA), هو (hw), ك (k), ني (ny), لهم (lhm), لكم (lkm), نا (nA), لنا (lnA), لها (lhA), لو (lw), لك (lk), and لي (ly). However, we also consider the concatenation of the above prefixes /suffixes (which is used in Arabic and its dialects). Hence, we also consider prefixes such as نت (nt), يت (yt) and suffixes such as هالكم (hAlkm) or هالهم (hAlhm).

4.1.2.4 Negation handling Negation in Arabic can be expressed with different words (لا, ليس, etc), while in its dialects is usually expressed by attaching a prefix, a suffix, or a combination of both. For example, the word مانحبكمش (*I don't like you*) can also be written as ما نحبكمش, مانحبكم ش or ما نحبكم ش. Hence, negation can be attached to or be separated from words. This work deals with both agglutinated and separated negation markers. A list of prefixes and suffixes related to negation are defined (including لا (lA), ما (mA), بلا (blA), مشى (mshy), ماشى (mAshy), ش (sh)). Different spaces were also concatenated to all the used prefixes / suffixes, as the user on social medias could use both writing (using / without using space between the prefixes/suffixes and the word). All the above prefixes /suffixes presented in section 4.1.2.3. are also used. The above affixes can be attached (concatenated) to the negation prefixes / suffixes. However, from the analysis of social media messages it can be seen that, in most cases, negation does not only affect the preceding word but also some of the words in the rest of the sentence. Thus, once a negation prefix or negation suffix is detected, the score of words following the negation is reversed, i.e. it is multiplied by -1.

4.2 Corpus-based approach

4.2.1 Corpus extraction and preprocessing

Text messages written in Arabic were extracted from the most popular Facebook pages used in Arabic countries[7]: MustafaHosny for Egypt with 32,854,861 fans[8]; ooredooqatar for Qatar with 834,031 fans[9] ; and EnnaharTv for Algeria with 9,603,348 fans[10]. In addition, the Facebook Rest API was used. To extract Arabic words, we use publicly available dictionaries, monolingual and parallel corpora [78,35,86]. To handle unstructured text, a set of preprocessing methods are used: (1) deletion of repeated messages; (2) removal of exaggerations; for example the word نحّحّحب is transformed into نحب (3) deletion of the character '#' and punctuation '.,!,?'; (4) removal of consecutive white spaces as well as the Arabic Tatweel ('–').

4.2.2 Corpus annotation

The lexicon constructed, as described in Section 4.1, is used to automatically assign a sentiment score to messages in the large corpus extracted from Face-

[7] https://www.socialbakers.com/statistics/facebook/pages/total/
[8] https://www.facebook.com/MustafaHosny/
[9] https://www.facebook.com/ooredooqatar/
[10] https://www.facebook.com/EnnaharTv/

book. Although our lexicon might be limited, messages that we automatically tagged as having strong positive or strong negative polarity may have other sentiment words. As such, we hope that subsequent training of a sentiment classifier on the automatically created corpus would improve results. Since we are interested in sentences that show strong sentiment, we retained sentences with a sentiment score above a threshold value α (as positives sentences) and sentences that have a sentiment score below the threshold value $-\alpha$ (as negatives sentences), and we varied α between 0 and 1 to determine its optimal value.

To increase labelling precision, the following heuristic approach was used:

1. If there are more positive sentiment words than negative sentiment words, then the message is considered positive (and vice versa). The role of this feature is to increase precision. Having a score greater/less than 0 does not certainly imply that the message is positive /negative. For proposing a corpus annotated automatically, we need to keep only the more precise samples. Hence, If the number of positive and negative sentiment words are equal, then we do not label the message.
2. The number of positive/negative sentiment words has to constitute at least 25% of the words in the message. This feature means that at least 1 word on 4 has to be in the lexicon. Then, if the message contains 8 words, 2 of them, have to be in the sentiment lexicon. The role of this feature is to avoid the case of a long message with only one word in the lexicon. Then, the found word could not be representative of the global sentiment of the message. For choosing 25%, other experiments were considered respectively using 50%, 75% and 100%. However, the most acceptable results were with 25%.

N	Sentences	Translation	Morphology handling	Features	Score calculation	Valence
1	نحب هاد العبة مليحة	I like this game, it's good	ن+حب هاد ا+لعبة مليح+ة	3 positive words	(+1.49)+(+1.22)+(2.09)/4=1.2	P
2	مليحة تهبل	It's good, it's make crazy	مليح+ة ت+هبل	1 positive word 1 negative word	(+2.09)+(-1.43)/2=0.33	NA
3	مليحة بصح تهبل	It's good **but** it's make crazy	مليح+ة بصح ت+هبل	Opposition	(-1.43)/1=-1.43	N
4	نموت على لفوت	I could die for the football meaning: I really like football	نموت على + لفوت	1 strong expression (نموت على)	+1.0	P
5	الفقر ماشي عيب وتاني ماشي مليح	Poorness isn't shameful and also it isn't good	ال+فقر ماشي عيب وتاني ماشي مليح	1 positive word 2 negation words (ماشي)	(-2.57)+(+2.17)+(-2.09) = -2.64/6=-0.44	N

Table 2 A sample of the automatic annotated corpus

For better clarifying the annotation process, we present in Table 2 five examples of annotated sentences (P–positive; N–negative) using the constructed lexicon and the presented annotation algorithm. If none of the heuristics rules presented above applies, then the message is not annotated and it is not added to the corpus (this case is presented by NA in the table). In order to present the process step by step, we also include Table 3 containing a small sample of the constructed lexicon with 6 terms, their translation, valence and intensity.

Terms	Translation	Valence	Intensity
حب	Love	P	+1.49
لعب	Play	P	+1.22
مليح	Good	P	+2.09
هبل	Making crazy	N	-1.43
فقر	Poorness	N	-2.57
عيب	Shameful	N	-2.17

Table 3 A sample of the constructed sentiment lexicon

The first entry message of Table 2 has a calculated sentiment score greater than 0 and the following heuristics rules apply: the number of positives words is greater that the number of negative words; the number of positives words represents 75% (3/4) of the number of words in the message. The second entry message, although there is a positive score recorded, is not annotated because the number of positive words equals the number of negative ones. The third entry message introduces the word "but" to entry message two, which expresses opposition. As explained before, the algorithm focuses only on the part following the opposition word ‘but’, which has only one word with negative valence and hence the message is annotated as negative. It can be seen that the fourth entry message contains the strong positive expression نموت على and therefore it is annotated as positive. For the last entry message in the table, the negation word ماشي inverses the polarity of the words; hence, it is annotated as negative.

4.2.3 Classification

Shallow and deep algorithms are used for classification. Features are extracted with word embedding techniques: Word2vec and Doc2vec algorithms are employed with shallow classification, while the methods embedding layer and fastText are employed with deep classification.

4.2.3.1 Word2vec + Machine learning (ML) Algorithms The model presented in [14] is used. In contrast to this work, we used both CBOW and SG representations and carried out a comparison between them.

4.2.3.2 fastText + Deep learning Algorithms The following model with five consecutive layers presented in [19] is used:

Layer1. This is a randomly-initialised word embedding layer that turns words in sentences into a feature map and preserves the spatial (contextual) information for each word.

Layer2. The feature map of Layer1 is scanned by a CNN.

Layer3. Global maxpooling is applied to the output generated by Layer2 to take the maximum score of each pattern. The main function of this pooling layer is to reduce the dimension of the CNN representations by down-sampling the output and keeping the maximum value.

Layer4. The scores from Layer3 are fed to a single feed-forward fully-connected layer with Relu activation.

Layer5. The output of Layer4 goes finally through a Softmax layer that predicts the output classes.

To enrich this model, herein the Continuous Bag of Words (CBOW) and Skip-Gram (SG) of FastText are used to compute the weights of embedding_matrix. In addition, the deep learning algorithms LSTM and Bi-LSTM are also used with the same CNN architecture.

5 Experimentation and simulations

The proposed Arabic SA approach is applied to the Algerian Maghrebi dialect (DALG), which is primarily used in informal communication including social media [67]. DALG is not used in school education or television news but in Algerian everyday life, music, etc., and it also goes by the names دارجة (*daArjah*), جزايري (*dziyriy*).

The experimental setup is presented below. The lexicon-based and corpus-based results are evaluated using precision, recall and F1-score metrics. Finally, an error analysis and corpus validation study are included.

5.1 Experimental setup

5.1.1 Data – lexicon and corpus

The English sentiment lexicon SOCAL [113], which contains 6,769 terms (2,827 adjectives; 1,039 adverbs; 1,761 nouns; 1,142 verbs), is used. SOCAL associates to each term its global sentiment score (-5 to -1 for negative terms; +1 to +5 for positive terms). Using Glosbe API, 3,952 of the 6,769 terms of SOCAL were recognised and translated, which resulted in an automatic Arabic sentiment lexicon ALGLex_V1 with 2,384 entries after associating average sentiment scores of repeated terms. The manual removal of ambiguous words led to ALGLex_V2 with 1,745 terms of which 968 are negative, 771 are positive and 6 are neutral.

We also constructed and used a set of corpora that are presented in more detail in the following:

i) A large Arabic corpus of 15,407,910 messages from the 226 most famous Algerian Facebook pages was created (November 2017) using RestFB[11]. This corpus contains 7,926,504 written in Arabic characters (Ar_corpus1).

[11] http://restfb.com/

ii) An annotated sentiment corpus, ALG_Senti, was automatically constructed based on ALGLex_V1 and a sentiment algorithm. This corpus contains 255,008 messages of which 127,004 are positive and 127,004 messages are negative. This corpus is very diverse and contains many subjects such as sport, politics, religions, products company, etc.

iii) Two test corpora for validating our automatic construction were created. The first, Senti_Alg_test, was created and manually annotated in [58] and contains 1,000 messages in Algerian dialect (500 in Arabic; 500 in Arabizi). In the context of this study we only focus on the Arabic part of this corpus. The second one, SANA_Alg, was created and manually annotated in [103] and contains 513 messages (236 positives; 194 negatives; 83 neutral) extracted from news, political, religion, sports, and society articles selected at the following Algerian Arabic newspaper web sites: Echorouk[12], Elkhabar[13], and Ennahar[14]. For comparison purposes, in order to proceed to a binary classification only on the positives and negatives messages (430 messages) are used.

5.1.2 Models

For the Word2vec model, we used the Gensim toolkit[15]. A context of 10 words was also used to produce representations for both CBOW, SG, PV-DBOW and PV-DM of length 300. The Word2vec model was trained on the Ar_corpus1. For the classification model, the implementation developed in [14] was used. This implementation calls Word2vec representation and five different classification algorithms: GaussianNB (GNB), LogisticRegression (LR), RandomForset (RF), SGDClassifier (SGD with loss = 'log' and penalty = 'l1') and LinearSVC (LSVC with C= '1e1'). For the deep learning algorithms, the implementation developed in [19] was used. This implementation is relying on 300 filters and a width of 7, i.e each filter is trained to detect a certain pattern in a 7-gram window of words. In order to classify sentiment in different languages, Attia et. [19], used a CNN model. In our presented work, four models with the same architecture are used: CNN, LSTM, MLP and Bi-LSTM. The embedding vectors are constructed using FastText, with the CBOW and SG representations using the same parameters as with Word2vec (i.e. window = 10 and size = 300). The following settings were used: epoch 100 with early stopping enabled. This parameter allows us to stop the models at an average of 20 epochs. The Adam optimiser was used for all the models.

[12] www.echoroukonline.com/ara/
[13] www.elkhabar.com
[14] www.ennaharonline.com
[15] https://radimrehurek.com/gensim/apiref.html

5.1.3 Metrics

In the context of this study, we use three metrics (Precision(P), Recall(R) and F1-score (F1)) for evaluating our sentiment analysis approach. Precision, as shown in Eq. 1, represents the number of sentiments correctly labelled as belonging to the positive class divided by the total number of sentiments labelled as belonging to the positive class. Recall, as shown in Eq. 2, represents the number of true positives divided by the total number of opinions that belongs to the positive class. Finally, F-score, as shown in Eq. 3, represents the harmonic mean of precision and recall [72].

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$F1 = \frac{2 * P * R}{P + R} = \frac{2 * TP}{2 * TP + FP + FN} \tag{3}$$

Where TP represents True Positive (i.e. manually annotated as positive and predicted by the model as positive). TN represents True Negative (i.e. manually annotated as negative and predicted by the model as negative). FP represents False Positive (i.e. manually annotated as negative and predicted by the model as positive). And FN represents False Negative (i.e. manually annotated as positive and predicted by the model by negative.

5.2 Experimental results

5.2.1 Lexicon-based approach results

Three metrics were used to evaluate the proposed Arabic SA approach: precision (P), recall (R) and F1 Score (F1). Table 4 shoes the different (P, R, F1) results obtained after the application of the proposed Arabic SA approach to the test corpora set: Senti (used in [58]) and SANA_Alg (used in [103]). The results are presented in relation to the two versions of the Algerian sentiment lexicon: the automatic constructed one (ALGLex_V1) and the manually reviewed one (ALGLex_V2). It is noticed that the ALGLex_V2 F1-score is the same as the ALGLex_V1 F1-score for Senti_Alg, while the ALGLex_V2 F1-score is slightly higher than the ALGLex_V1 F1-score in the case of SANA_Alg. The manual review involved in the construction of ALGLex_V2 is reflected in its precision and recall being higher than and lower than the ALGLex_V1 precision and recall, respectively, for both test corpora. The automatic annotation needs to be as precise as possible, hence the increase in precision of our proposed lexicon-based approach fits perfectly with our ultimate goal.

Lexicon Version	Senti_Alg			SANA_Alg		
	P	R	F1	P	R	F1
ALGLex_V1	0.76	0.68	0.72	0.54	0.48	0.51
ALGLex_V2	0.81	0.65	0.72	0.58	0.47	0.52

Table 4 Lexicon-based approach applied on Algerian Dialect

5.2.2 Corpus-based approach results

The construction of our annotated corpus ALG_Senti is firstly based on the sentiment score returned by our lexicon-based approach. In the context of this study, we consider that a message is potentially positive if its sentiment score is greater than 0, and potentially negative when its sentiment score is lower than 0. However, other features for annotation are considered as we aim to increase the precision of our annotation process. We denote the feature related to score by *Score*, the feature related to the number of positive/negative words related to the total number of words by *Compare_total*, and the feature related to comparison between the number of positive/negative words by *Compare_pos_neg*. The importance of using these features are shown in Table 5, which illustrates the results obtained when implementing each feature. It can be seen that the application of each feature increases the precision, from 81% to 89% (for Senti_Alg) and from 58% to 100% (for SANA_Alg). The increase of precision is related to the decrease of recall. However, in our context, precision is more important than recall. This is because we start with a very large and voluminous Arabic corpus and, even with a minimum recall, the resulted annotated corpus is still voluminous.

Table 6 presents the proposed system implemented with Word2vec + ML algorithms and FastText with deep learning algorithms on Senti_Alg and SANA_Alg, respectively. Regarding the classification algorithms, Table 6 results are mitigated between SG and CBOW models, i.e. with some ML classifiers, such as GNB, CBOW outperforms SG on both test corpora, while with others, such as LSVC, SG outperforms CBOW. The same observation could be drawn with deep learning classifiers: CBOW outperforms SG with CNN, while SG outperforms CBOW with MLP. Regarding the ML classifiers, the SGD classifier outperforms the other classifiers with both CBOW and SG models on Senti_Alg with an F1 up to 87.77% with CBOW and up to 86.27% on SG; while the GNB outperforms the other classifiers with both CBOW and SG models on SANA_Alg with an F1 up to 81.00% with CBOW and up to 75.82% with SG. For deep learning classifiers, it can be seen that LSTM and MLP outperform the other classifiers with the best F1 (80.40%) on Senti_Alg being achieved by SG model with the MLP classifier while on SANA_Alg the best F1 (61.99%) is achieved using CBOW with LSTM. To sum up, using Word2vec with shallow classifiers outperform fastText with deep learning classifiers on both test corpora Senti_Alg and SANA_Alg. This is perfectly

AGLex_V2	Senti_Alg			SANA_Alg		
	P	R	F1	P	R	F1
Score	0.81	0.65	0.72	0.58	0.47	0.52
Compare_total	0.88	0.29	0.43	1.0	0.01	0.02
Compare_pos_neg	0.89	0.28	0.42	1.0	0.01	0.02

Table 5 Annotation process results by applying different features using ALGLex_V2

Model	Type	Classif. alg.	Senti_Alg			SANA_Alg		
			P	R	F1	P	R	F1
Word2vec	CBOW	GNB	93.50	74.80	83.11	81.17	80.23	81.00
		LR	82.09	88.00	84.94	76.23	70.83	73.43
		RF	85.07	75.20	79.83	71.54	77.50	74.40
		SGD	85.28	90.40	87.77	80.28	72.92	76.42
		LSVC	82.71	88.00	85.27	74.44	69.17	71.71
	SG	GNB	90.34	74.80	81.84	62.37	96.67	75.82
		LR	85.10	86.80	85.94	79.09	72.50	75.65
		RF	85.59	76.00	80.51	72.05	76.25	74.09
		SGD	84.62	88.00	86.27	82.74	67.92	74.60
		LSVC	85.32	86.00	85.66	78.80	71.25	74.84
FastText	CBOW	CNN	78.06	78.00	77.99	59.34	59.77	59.20
		MLP	78.00	78.00	78.00	57.57	57.93	57.62
		LSTM	80.24	80.20	80.19	61.95	62.07	61.99
		Bi-LSTM	80.03	80.00	79.99	61.00	61.15	61.05
	SG	CNN	80.33	80.00	79.95	58.27	58.85	57.66
		MLP	80.41	80.40	80.40	58.84	59.08	58.90
		LSTM	79.00	79.00	79.00	59.33	59.54	59.39
		Bi-LSTM	77.61	77.60	77.60	59.56	60.00	58.59

Table 6 Shallow and deep classification results on the corpus constructed automatically

understandable because deep learning is more adequate for data annotated manually and with higher annotation precision.

5.3 Discussion and Error Analysis

Our Arabic SA approach is based on an automatic corpus annotation with a sentiment lexicon that has been applied to the Algerian dialect. In order to compare our results with those presented in the literature, we use the same test corpora used in both [58] and [103]. The results for Arabic presented in [58] had an F1 up to 68% while the application of our Arabic SA approach has an F1 up to 88%, i.e. an improvement of 20 percentage points. The results presented in [103] had an F1 up to 75% while the application of our Arabic SA approach has an F1 up to 81%, i.e. an improvement of 6 percentage points. However, it is worth noting the following issues:

– The quality of the sentiment lexicon definitely affects the quality of the automatic annotation. The following drawbacks were observed with lexicon construction and with the proposed algorithm for sentiment score computation:

- Irregular plural. In general, in Arabic and its dialects, the plural is formed by adding the same suffixes; although there are some words that do not follow the plural regular forms. For example the plural of the word مليح (*good*) is neither the regular form مليحين nor the regular form مليهون, but rather ملاح.
- The non-presence of certain words in the lexicon. Certain words, like كادو (*a gift*), are not present in our lexicon; hence they could not be considered for sentiment scores.
- The non handling of intensifiers. Certain adjective such as بزاف (*very*) intensify the sentiment of words.

Thus, it is crucial to handle them properly for improving the computation of sentiment.

- The most important classification errors were due to some errors that occurred in the automatic annotated corpus, i.e. in the training corpus construction. For example the messages جابو فخامة الاسم تكفي (*Djabou the excellency of the name is sufficient*) was annotated negative when it is positive. Another example is the message يهدي العب اله جيب الخير (*guide the play, we hope god bring the good things*) was wrongly annotated as positive when it is known that "we hope god bring the good things" is an expression used to speak about bad things. Thus, manually reviewing the automatic annotation will definitely improve the results.

5.4 Corpus validation

To validate the constructed corpus automatically, we focus on a sample containing 3,048 messages (1,488 positives ; 1,560 negatives). Afterwards, we manually review this sample. The messages that are correctly annotated are kept and those which are wrongly annotated are corrected. In addition, some objective messages not possessing a sentiment are deleted. Our first observation is that, among the 3,048 messages that are manually reviewed, 85.17% (2,596 messages) are correctly annotated. To the best of our knowledge, this corpus is the first manually checked annotated sentiment corpus that handles DALG as well as MSA.

The utility of the manual reviewing is shown in Table 7 results of classification after a manual review of the corpus. It can be seen clearly that the manual reviewing of the corpus improve slightly the F1 on Senti_Alg: up to 90% using Word2vec SG with SGD classifier. No improvement on SANA_Alg was observed with shallow classifiers where the best result F1 remained 81%. However, by using the manually reviewed corpus, the results obtained with deep learning classifiers are drastically improved: the best F1 on Senti_Alg is 90.20% using FastText SG with MLP classifier; the best F1 on SANA_Alg using the corpus constructed automatically was is 61.99%, which increased after the manual review to 71.56%, i.e. an improvement of nearly 10 points. As it was mentioned earlier, deep learning classifiers give better results the higher the data accuracy is, although they require large datasets. A corpus containing 3,048 messages is clearly not large enough to obtain good results using deep learning.

Model	Type	Classif. alg.	Senti_Alg			SANA_Alg		
			P	R	F1	P	R	F1
Word2vec	CBOW	GNB	93.65	70.80	80.64	77.73	68.33	72.73
		LR	94.50	82.40	88.03	87.94	72.92	79.73
		RF	85.78	70.00	77.09	85.16	55.00	66.84
		SGD	86.97	90.80	88.85	91.36	61.67	73.63
		LSVC	89.58	86.00	87.76	90.00	63.75	74.63
	SG	GNB	93.09	70.00	79.91	66.77	93.75	77.99
		LR	95.43	83.60	89.13	86.32	76.25	80.97
		RF	91.58	69.60	79.09	85.92	50.83	63.87
		SGD	88.76	91.60	90.16	83.41	75.42	79.21
		LSVC	91.77	84.80	88.15	84.32	65.00	73.41
FastText	CBOW	CNN	87.54	87.40	87.39	73.04	71.26	71.23
		MLP	86.33	86.20	86.19	70.30	70.11	70.17
		LSTM	86.21	86.20	86.20	72.21	71.26	71.32
		Bi-LSTM	88.27	88.20	88.19	74.98	71.95	71.75
	SG	CNN	88.09	88.00	87.99	70.17	69.66	69.74
		MLP	90.28	90.20	90.20	70.55	70.57	70.56
		LSTM	86.86	86.60	86.58	71.62	69.89	69.85
		Bi-LSTM	88.04	88.00	88.00	72.40	71.49	71.56

Table 7 Shallow and deep classification results on the validated corpus

6 Conclusion and perspective

In this paper, we presented a robust approach for SA of Arabic and its dialects. To do this, we firstly created an Arabic sentiment lexicon based on a translated English lexicon. We employed the constructed lexicon in the creation of a large automatically tagged sentiment corpus of Algerian Facebook messages that were automatically extracted using Facebook RestAPI. Further, we focused on handling the morphological characteristics of Arabic and its dialects. For classification we used shallow (GNB, LR, RF, SGD, LSVC) and deep (CNN, MLP, LSTM, Bi-LSTM) classifiers. For shallow classification we used Word2vec while for deep classification we used fastText. For validating and comparing the constructed corpus, we carried out a set of experiments on two external tests set, and it was observed that our approach outperforms the results presented in the research literature. We also focused on a set of samples that we manually reviewed and it was noted that 85.17% were correctly annotated. Although this approach was applied to an under-resource Arabic language, it is obvious that it was a generic approach that can be extended to other languages. In addition, the same methodology could be utilised in other NLP tasks that require annotated data.

There are issues when applying the proposed Arabic SA approach. The most important one is the lack of precision in the automatic annotation, which we addressed with a manual review. However, to answer the other drawbacks, we plan to focus in future on the following aspects:

1. To enrich the proposed lexicon with calling Word2vec. The aim of Word2vec is to return the semantically closest word to a given word (i.e. words with similar vectors). However, the problem with this technique is that the two words 'good' and 'bad' are returned as too close. These two words appear frequently in the same context. Hence it is crucial to resolve this 'good/bad' issue.

2. We observed that the manually reviewed corpus improved slightly the results obtained when shallow classifiers were used while the improvements were more significant when deep learning classifiers were used. However, the small size of the corpus limited these improvements, and therefore increasing the size of the manually reviewed corpus would improve the results. In any case, reviewing an automatically constructed corpus would be certainly less consuming, in terms of time and effort required, than constructing such a corpus from scratch.

3. In this paper, we focused on two major aspects of automatic annotations: emoticons and sentiment lexicon. However, other directions could have been followed such as the one based on prediction. Hence, we plan to construct an annotated corpus (manually or semi-automatically) with a limited number of messages (1000 messages), as general as possible so as to handle as many topics and domains as possible, to generate a model (shallow; deep; hybrid) to predict and automatically annotate other messages extracted from a voluminous corpus.

4. Other prominent issues to investigate include the non-standard romanisation (called Arabizi) that Arabic speakers often use in social media. Arabizi uses Latin alphabet, numbers, punctuation for writing Arabic words: for example, the word 'mli7' is the romanised form of the Arabic word 'مليح' ('*good*'). Recent work has been carried out for handling Arabizi in [58,57]. In the first one, an Arabizi sentiment corpus was constructed automatically while a transliteration step was presented for handling Arabizi in the second one, which relatively improved the results with a rule-based transliteration approach. Our future aim includes the proposal of an approach for handling Arabic and Arabizi, and to develop a statistical approach for Arabizi transliteration.

5. In future, we also plan to automatically construct resources (lexicon and corpus) for dialects other than Algerian (Tunisian, Moroccan, Egyptian, etc.). It would also be interesting to apply this approach to languages such as English or Chinese.
6. Finally, we aim to extend the proposed approach to other NLP tasks such as identification, hate detection, fake news detection, etc.

To sum up, this paper presented and validated an automatically constructed corpus, dedicated to Arabic sentiment analysis. The sentiment classification was done using both shallow and deep algorithm. However, the encouraging results obtained by using this approach lead to some open issues principally related to the generalisation. Indeed, this approach could be generalised to other dialects, other languages and also to other fields. Hence, the aim of our future work is to carry out and to analyse the generalisation of this approach.

References

1. Abdul-Mageed, M., Diab, M.: Toward building a large-scale arabic sentiment lexicon. In: Proceedings of the 6th international global WordNet conference, pp. 18–22 (2012)
2. Abdul-Mageed, M., Diab, M., Kübler, S.: Samar: Subjectivity and sentiment analysis for arabic social media. Computer Speech & Language **28**(1), 20–37 (2014)
3. Abdul-Mageed, M., Diab, M.T.: Awatif: A multi-genre corpus for modern standard arabic subjectivity and sentiment analysis. In: LREC, pp. 3907–3914. Citeseer (2012)
4. Abdul-Mageed, M., Diab, M.T.: Sana: A large scale multi-genre, multi-dialect lexicon for arabic subjectivity and sentiment analysis. In: LREC (2016)
5. Abdulla, N., Mohammed, S., Al-Ayyoub, M., Al-Kabi, M., et al.: Automatic lexicon construction for arabic sentiment analysis. In: Future Internet of Things and Cloud (FiCloud), 2014 International Conference on, pp. 547–552. IEEE (2014)
6. Abdulla, N.A., Ahmed, N.A., Shehab, M.A., Al-Ayyoub, M.: Arabic sentiment analysis: Lexicon-based and corpus-based. In: Applied Electrical Engineering and Computing Technologies (AEECT), 2013 IEEE Jordan Conference on, pp. 1–6. IEEE (2013)
7. Abdulla, N.A., Ahmed, N.A., Shehab, M.A., Al-Ayyoub, M., Al-Kabi, M.N., Al-rifai, S.: Towards improving the lexicon-based approach for arabic sentiment analysis. International Journal of Information Technology and Web Engineering (IJITWE) **9**(3), 55–71 (2014)
8. Al-Azani, S., El-Alfy, E.S.M.: Using word embedding and ensemble learning for highly imbalanced data sentiment analysis in short arabic text. Procedia Computer Science **109**, 359–366 (2017)
9. Al-Sallab, A., Baly, R., Hajj, H., Shaban, K.B., El-Hajj, W., Badaro, G.: Aroma: a recursive deep learning model for opinion mining in arabic as a low resource language. ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) **16**(4), 25 (2017)
10. Al Shboul, B., Al-Ayyoub, M., Jararweh, Y.: Multi-way sentiment classification of arabic reviews. In: 2015 6th International Conference on Information and Communication Systems (ICICS), pp. 206–211. IEEE (2015)

11. Al-Twairesh, N., Al-Khalifa, H., Al-Salman, A., Al-Ohali, Y.: Arasenti-tweet: A corpus for arabic sentiment analysis of saudi tweets. Procedia Computer Science **117**, 63–72 (2017)

12. Alayba, A.M., Palade, V., England, M., Iqbal, R.: A combined cnn and lstm model for arabic sentiment analysis. In: International Cross-Domain Conference for Machine Learning and Knowledge Extraction, pp. 179–191. Springer (2018)

13. Alowaidi, S., Saleh, M., Abulnaja, O.: Semantic sentiment analysis of arabic texts. International Journal of Advanced Computer Science and Applications **8**(2), 256–262 (2017)

14. Altowayan, A.A., Tao, L.: Word embeddings for arabic sentiment analysis. In: Big Data (Big Data), 2016 IEEE International Conference on, pp. 3820–3825. IEEE (2016)

15. Altrabsheh, N., El-Masri, M., Mansour, H.: Combining sentiment lexicons of arabic terms (2017)

16. Aly, M., Atiya, A.: Labr: A large scale arabic book reviews dataset. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), vol. 2, pp. 494–498 (2013)

17. Arora, M., Kansal, V.: Character level embedding with deep convolutional neural network for text normalization of unstructured data for twitter sentiment analysis. Social Network Analysis and Mining **9**(1), 12 (2019)

18. Atia, S., Shaalan, K.: Increasing the accuracy of opinion mining in arabic. In: Arabic Computational Linguistics (ACLing), 2015 First International Conference on, pp. 106–113. IEEE (2015)

19. Attia, M., Samih, Y., El-Kahky, A., Kallmeyer, L.: Multilingual multi-class sentiment classification using convolutional neural networks. In: LREC (2018)

20. Baccianella, S., Esuli, A., Sebastiani, F.: Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: LREC, vol. 10

21. Badaro, G., Baly, R., Akel, R., Fayad, L., Khairallah, J., Hajj, H., Shaban, K., El-Hajj, W.: A light lexicon-based mobile application for sentiment mining of arabic tweets. In: Proceedings of the Second Workshop on Arabic Natural Language Processing, pp. 18–25 (2015)

22. Badaro, G., Baly, R., Hajj, H., Habash, N., El-Hajj, W.: A large scale arabic sentiment lexicon for arabic opinion mining. In: Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP), pp. 165–173 (2014)

23. Banea, C., Mihalcea, R., Wiebe, J.: Porting multilingual subjectivity resources across languages. IEEE Transactions on Affective Computing **4**(2), 211–225 (2013)

24. Barhoumi, A., Aloulou, Y.E.C., Belguith, L.H.: Document embeddings for arabic sentiment analysis (2017)

25. Bisio, F., Meda, C., Gastaldo, P., Zunino, R., Cambria, E.: Sentiment-oriented information retrieval: Affective analysis of documents based on the senticnet framework. In: Sentiment analysis and ontology engineering, pp. 175–197. Springer (2016)

26. Blitzer, J., Dredze, M., Pereira, F.: Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In: Proceedings of the 45th annual meeting of the association of computational linguistics, pp. 440–447 (2007)

27. Boudad, N., Faizi, R., Thami, R.O.H., Chiheb, R.: Sentiment analysis in arabic: A review of the literature. Ain Shams Engineering Journal (2017)

28. Buckwalter, T.: Buckwalter arabic morphological analyzer version 2.0. linguistic data consortium, university of pennsylvania, 2002. ldc cat alog no.: Ldc2004l02. Tech. rep., ISBN 1-58563-324-0 (2004)

29. Cambria, E., Hussain, A., Vinciarelli, A.: Affective reasoning for big social data analysis. IEEE Transactions on Affective Computing **8**(4), 426–427 (2017)

30. Cambria, E., Speer, R., Havasi, C., Hussain, A.: Senticnet: A publicly available semantic resource for opinion mining. In: AAAI fall symposium: commonsense knowledge, vol. 10 (2010)

31. Chen, H., Sun, M., Tu, C., Lin, Y., Liu, Z.: Neural sentiment classification with user and product attention. In: Proceedings of the 2016 conference on empirical methods in natural language processing, pp. 1650–1659 (2016)

32. Cramer, J.S.: The origins of logistic regression (2002)

33. Dahou, A., Xiong, S., Zhou, J., Haddoud, M.H., Duan, P.: Word embeddings and convolutional neural network for arabic sentiment classification. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pp. 2418–2427 (2016)

34. Denecke, K.: Using sentiwordnet for multilingual sentiment analysis. In: Data Engineering Workshop, 2008. ICDEW 2008. IEEE 24th International Conference on, pp. 507–512. IEEE (2008)

35. Diab, M.T., Al-Badrashiny, M., Aminian, M., Attia, M.: Tharwa: A large scale dialectal arabic-standard arabic-english lexicon. In: LREC (2014)

36. Ding, X., Liu, B., Yu, P.S.: A holistic lexicon-based approach to opinion mining. In: Proceedings of the 2008 international conference on web search and data mining, pp. 231–240. ACM (2008)

37. Dou, Z., Wei, W., Wan, X.: Improving word embeddings for antonym detection using thesauri and sentiwordnet. In: CCF International Conference on Natural Language Processing and Chinese Computing, pp. 67–79. Springer (2018)

38. Dou, Z.Y.: Capturing user and product information for document level sentiment analysis with deep memory network. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 521–526 (2017)

39. El-Beltagy, S.R.: Niletmrg at semeval-2016 task 7: Deriving prior polarities for arabic sentiment terms. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), pp. 486–490 (2016)

40. El-Beltagy, S.R.: Nileulex: A phrase and word level sentiment lexicon for egyptian and modern standard arabic. In: LREC (2016)

41. El-Beltagy, S.R.: Weightednileulex: A scored arabic sentiment lexicon for improved sentiment analysis. Language Processing, Pattern Recognition and Intelligent Systems. Special Issue on Computational Linguistics, Speech& Image Processing for Arabic Language. World Scientific Publishing Co (2017)

42. El-Beltagy, S.R., Ali, A.: Open issues in the sentiment analysis of arabic social media: A case study. In: Innovations in information technology (iit), 2013 9th international conference on, pp. 215–220. IEEE (2013)

43. El-Kilany, A., Azzam, A., El-Beltagy, S.R.: Using deep neural networks for extracting sentiment targets in arabic tweets. In: Intelligent Natural Language Processing: Trends and Applications, pp. 3–15. Springer (2018)

44. El Mahdaouy, A., Gaussier, E., El Alaoui, S.O.: Arabic text classification based on word and document embeddings. In: International Conference on Advanced Intelligent Systems and Informatics, pp. 32–41. Springer (2016)

45. ElSahar, H., El-Beltagy, S.R.: A fully automated approach for arabic slang lexicon extraction from microblogs. In: International Conference on Intelligent Text Processing and Computational Linguistics, pp. 79–91. Springer (2014)

46. ElSahar, H., El-Beltagy, S.R.: Building large arabic multi-domain resources for sentiment analysis. In: International Conference on Intelligent Text Processing and Computational Linguistics, pp. 23–34. Springer (2015)

47. Eskander, R., Rambow, O.: Slsa: A sentiment lexicon for standard arabic. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 2545–2550 (2015)

48. Esuli, A., Sebastiani, F.: Sentiwordnet: a high-coverage lexical resource for opinion mining. Evaluation **17**, 1–26 (2007)

49. Farghaly, A., Shaalan, K.: Arabic natural language processing: Challenges and solutions. ACM Transactions on Asian Language Information Processing (TALIP) **8**(4), 14 (2009)

50. Farra, N., McKeown, K.: Smarties: Sentiment models for arabic target entities. arXiv preprint arXiv:1701.03434 (2017)

51. Fellbaum, C., Alkhalifa, M., Black, W., Elkateb, S., Pease, A., Rodriguez, H., Vossen, P.: Introducing the arabic wordnet project (2006)

52. Fukushima, K.: Neural network model for a mechanism of pattern recognition unaffected by shift in position-neocognitron. IEICE Technical Report, A **62**(10), 658–665 (1979)

53. Gamal, D., Alfonse, M., El-Horbaty, E.S.M., Salem, A.B.M.: Twitter benchmark dataset for arabic sentiment analysis. International Journal of Modern Education and Computer Science **11**(1), 33 (2019)

54. Gatti, L., Guerini, M., Turchi, M.: Sentiwords: Deriving a high precision and high coverage lexicon for sentiment analysis. IEEE Transactions on Affective Computing **7**(4), 409–421 (2016)

55. Gilbert, B., Hussein, J., Hazem, H., Wassim, E.H., Nizar, H.: Arsel: A large scale arabic sentiment and emotion lexicon (2018)

56. Graff, D., Buckwalter, T., Jin, H., Maamouri, M.: Lexicon development for varieties of spoken colloquial arabic. In: LREC (2006)

57. Guellil, I., Adeel, A., Azouaou, F., Benali, F., Hachani, A.e., Hussain, A.: Arabizi sentiment analysis based on transliteration and automatic corpus annotation. In: Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp. 335–341 (2018)

58. Guellil, I., Adeel, A., Azouaou, F., Hussain, A.: Sentialg: Automated corpus annotation for algerian sentiment analysis. arXiv preprint arXiv:1808.05079 (2018)

59. Guellil, I., Azouaou, F.: Arabic dialect identification with an unsupervised learning (based on a lexicon). application case: Algerian dialect. In: Computational Science and Engineering (CSE) and IEEE Intl Conference on Embedded and Ubiquitous Computing (EUC) and 15th Intl Symposium on Distributed Computing and Applications for Business Engineering (DCABES), 2016 IEEE Intl Conference on, pp. 724–731. IEEE (2016)

60. Guellil, I., Azouaou, F., Mendoza, M.: Arabic sentiment analysis: studies, resources, and tools. Social Network Analysis and Mining **9**(1), 56 (2019)

61. Guellil, I., Azouaou, F., Saâdane, H., Semmar, N.: Une approche fondée sur les lexiques d'analyse de sentiments du dialecte algérien (2017)

62. Guellil, I., Azouaou, F., Valitutti, A.: English vs arabic sentiment analysis: A survey presenting 100 work studies, resources and tools. In: 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA), pp. 1–8. IEEE (2019)

63. Guellil, I., Boukhalfa, K.: Social big data mining: A survey focused on opinion mining and sentiments analysis. In: Programming and Systems (ISPS), 2015 12th International Symposium on, pp. 1–10. IEEE (2015)

64. Habash, N.Y.: Introduction to arabic natural language processing. Synthesis Lectures on Human Language Technologies **3**(1), 1–187 (2010)

65. Hamouda, A., Rohaim, M.: Reviews classification using sentiwordnet lexicon. In: World congress on computer science and information technology. sn (2011)

66. Harrat, S., Meftouh, K., Abbas, M., Smaili, K.: Building resources for algerian arabic dialects. In: Fifteenth Annual Conference of the International Speech Communication Association (2014)

67. Harrat, S., Meftouh, K., Smaïli, K.: Machine translation for arabic dialects (survey). Information Processing & Management (2017)

68. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**(8), 1735–1780 (1997)

69. Hogenboom, A., Bal, D., Frasincar, F., Bal, M., de Jong, F., Kaymak, U.: Exploiting emoticons in sentiment analysis. In: Proceedings of the 28th Annual ACM Symposium on Applied Computing, pp. 703–710. ACM (2013)

70. Htait, A., Fournier, S., Bellot, P.: Lsis at semeval-2017 task 4: Using adapted sentiment similarity seed words for english and arabic tweet polarity classification. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), pp. 718–722 (2017)

71. Joulin, A., Grave, E., Bojanowski, P., Mikolov, T.: Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759 (2016)

72. Keyvanpour, M., Zandian, Z.K., Heidarypanah, M.: Omlml: a helpful opinion mining method based on lexicon and machine learning in social networks. Social Network Analysis and Mining **10**(1), 1–17 (2020)

73. Khan, F.H., Qamar, U., Bashir, S.: A semi-supervised approach to sentiment analysis using revised sentiment strength based on sentiwordnet. Knowledge and information Systems **51**(3), 851–872 (2017)

74. Khoja, S., Garside, R.: Stemming arabic text. Lancaster, UK, Computing Department, Lancaster University (1999)

75. Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882 (2014)

76. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

77. Kumar, A., Kohail, S., Kumar, A., Ekbal, A., Biemann, C.: Iit-tuda at semeval-2016 task 5: Beyond sentiment lexicon: Combining domain dependency and distributional semantics features for aspect based sentiment analysis. In: Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016), pp. 1129–1135 (2016)

78. Kwaik, K.A., Saad, M., Chatzikyriakidis, S., Dobnik, S.: Shami: A corpus of levantine arabic dialects (2018)

79. Le, Q., Mikolov, T.: Distributed representations of sentences and documents. In: International Conference on Machine Learning, pp. 1188–1196 (2014)

80. Liu, B.: Sentiment analysis and opinion mining. Synthesis lectures on human language technologies **5**(1), 1–167 (2012)

81. Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y., Potts, C.: Learning word vectors for sentiment analysis. In: Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1, pp. 142–150. Association for Computational Linguistics (2011)

82. Mahyoub, F.H., Siddiqui, M.A., Dahab, M.Y.: Building an arabic sentiment lexicon using semi-supervised learning. Journal of King Saud University-Computer and Information Sciences **26**(4), 417–424 (2014)

83. Mataoui, M., Zelmati, O., Boumechache, M.: A proposed lexicon-based sentiment analysis approach for the vernacular algerian arabic. Research in Computing Science **110**, 55–70 (2016)

84. McAuley, J., Leskovec, J.: Hidden factors and hidden topics: understanding rating dimensions with review text. In: Proceedings of the 7th ACM conference on Recommender systems, pp. 165–172. ACM (2013)

85. Medhaffar, S., Bougares, F., Esteve, Y., Hadrich-Belguith, L.: Sentiment analysis of tunisian dialects: Linguistic ressources and experiments. In: Proceedings of the Third Arabic Natural Language Processing Workshop, pp. 55–61 (2017)

86. Meftouh, K., Harrat, S., Jamoussi, S., Abbas, M., Smaili, K.: Machine translation experiments on padic: A parallel arabic dialect corpus. In: The 29th Pacific Asia conference on language, information and computation (2015)

87. Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., Freeman, J., Tsai, D., Amde, M., Owen, S., et al.: Mllib: Machine learning in apache spark. The Journal of Machine Learning Research **17**(1), 1235–1241 (2016)

88. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)

89. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems, pp. 3111–3119 (2013)

90. Miller, G.A.: Wordnet: a lexical database for english. Communications of the ACM **38**(11), 39–41 (1995)

91. Mohammad, S., Salameh, M., Kiritchenko, S.: Sentiment lexicons for arabic social media. In: LREC (2016)

92. Mohammad, S.M., Salameh, M., Kiritchenko, S.: How translation alters sentiment. Journal of Artificial Intelligence Research **55**, 95–130 (2016)

93. Mohammed, A., Kora, R.: Deep learning approaches for arabic sentiment analysis. Social Network Analysis and Mining **9**(1), 52 (2019)

94. Mourad, A., Darwish, K.: Subjectivity and sentiment analysis of modern standard arabic and arabic microblogs. In: Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis, pp. 55–64 (2013)

95. Nabil, M., Aly, M., Atiya, A.: Astd: Arabic sentiment tweets dataset. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 2515–2519 (2015)

96. Nagamanjula, R., Pethalakshmi, A.: A novel framework based on bi-objective optimization and lan2fis for twitter sentiment analysis. Social Network Analysis and Mining **10**(1), 34 (2020)

97. Oghina, A., Breuss, M., Tsagkias, M., de Rijke, M.: Predicting imdb movie ratings using social media. In: European Conference on Information Retrieval, pp. 503–507. Springer (2012)

98. Pak, A., Paroubek, P.: Twitter as a corpus for sentiment analysis and opinion mining. (2010)

99. Panos, A., Dellaportas, P., Titsias, M.K.: Fully scalable gaussian processes using subspace inducing inputs. arXiv preprint arXiv:1807.02537 (2018)

100. Poria, S., Gelbukh, A., Cambria, E., Yang, P., Hussain, A., Durrani, T.: Merging senticnet and wordnet-affect emotion lists for sentiment analysis. In: Signal Processing (ICSP), 2012 IEEE 11th International Conference on, vol. 2, pp. 1251–1255. IEEE (2012)

101. Poria, S., Gelbukh, A., Hussain, A., Howard, N., Das, D., Bandyopadhyay, S.: Enhanced senticnet with affective labels for concept-based opinion mining. IEEE Intelligent Systems **28**, 31–38 (2013)

102. Rahab, H., Zitouni, A., Djoudi, M.: Siaac: Sentiment polarity identification on arabic algerian newspaper comments. In: Proceedings of the Computational Methods in Systems and Software, pp. 139–149. Springer (2017)

103. Rahab, H., Zitouni, A., Djoudi, M.: Sana: Sentiment analysis on newspapers comments in algeria. Journal of King Saud University-Computer and Information Sciences (2019)

104. Ren, F., Matsumoto, K.: Semi-automatic creation of youth slang corpus and its application to affective computing. IEEE Transactions on Affective Computing **7**(2), 176–189 (2016)

105. Rushdi-Saleh, M., Martín-Valdivia, M.T., Ureña-López, L.A., Perea-Ortega, J.M.: Bilingual experiments with an arabic-english corpus for opinion mining. In: Proceedings of the International Conference Recent Advances in Natural Language Processing 2011, pp. 740–745 (2011)

106. Rushdi-Saleh, M., Martín-Valdivia, M.T., Ureña-López, L.A., Perea-Ortega, J.M.: Oca: Opinion corpus for arabic. Journal of the Association for Information Science and Technology **62**(10), 2045–2054 (2011)

107. Saadane, H., Habash, N.: A conventional orthography for algerian arabic. In: ANLP Workshop 2015, p. 69 (2015)

108. Saadane, H., Seffih, H., Fluhr, C., Choukri, K., Semmar, N.: Automatic identification of maghreb dialects using a dictionary-based approach. In: LREC (2018)

109. Sadat, F., Mallek, F., Boudabous, M., Sellami, R., Farzindar, A.: Collaboratively constructed linguistic resources for language variants and their exploitation in nlp application–the case of tunisian arabic and the social media. In: Proceedings of Workshop on Lexical and Grammatical Resources for Language Processing, pp. 102–110 (2014)

110. Salameh, M., Mohammad, S., Kiritchenko, S.: Sentiment after translation: A case-study on arabic social media posts. In: Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: Human language technologies, pp. 767–777 (2015)

111. Schmidhuber, J.: Deep learning in neural networks: An overview. Neural networks **61**, 85–117 (2015)

112. Schmitt, M., Steinheber, S., Schreiber, K., Roth, B.: Joint aspect and polarity classification for aspect-based sentiment analysis with end-to-end neural networks. arXiv preprint arXiv:1808.09238 (2018)

113. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. Computational linguistics **37**(2), 267–307 (2011)

114. Tafreshi, S., Diab, M.: Emotion detection and classification in a multigenre corpus with joint multi-task deep learning. In: Proceedings of the 27th International Conference on Computational Linguistics, pp. 2905–2913 (2018)

115. Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., Qin, B.: Learning sentiment-specific word embedding for twitter sentiment classification. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol. 1, pp. 1555–1565 (2014)

116. Tellez, E.S., Miranda-Jiménez, S., Graff, M., Moctezuma, D., Suárez, R.R., Siordia, O.S.: A simple approach to multilingual polarity classification in twitter. Pattern Recognition Letters **94**, 68–74 (2017)

117. Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., Kappas, A.: Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology **61**(12), 2544–2558 (2010)

118. Tofighy, S., Fakhrahmad, S.M.: A proposed scheme for sentiment analysis: Effective feature reduction based on statistical information of sentiwordnet. Kybernetes **47**(5), 957–984 (2018)

119. Tomar, D.S., Sharma, P.: A text polarity analysis using sentiwordnet based an algorithm. IJCSIT) International Journal of Computer Science and Information Technologies (2016)

120. Vapnik, V.: Pattern recognition using generalized portrait method. Automation and remote control **24**, 774–780 (1963)

121. Vo, D.T., Zhang, Y.: Don't count, predict! an automatic approach to learning sentiment lexicons for short text. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), vol. 2, pp. 219–224 (2016)

122. Wiebe, J., Wilson, T., Cardie, C.: Annotating expressions of opinions and emotions in language. Language resources and evaluation **39**(2-3), 165–210 (2005)

123. Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., Patwardhan, S.: Opinionfinder: A system for subjectivity analysis. In: Proceedings of hlt/emnlp on interactive demonstrations, pp. 34–35. Association for Computational Linguistics (2005)

124. Xia, R., Jiang, J., He, H.: Distantly supervised lifelong learning for large-scale social media sentiment analysis. IEEE Transactions on Affective Computing **8**(4), 480–491 (2017)

125. Yadav, P., Pandya, D.: Sentireview: Sentiment analysis based on text and emoticons. In: Innovative Mechanisms for Industry Applications (ICIMIA), 2017 International Conference on, pp. 467–472. IEEE (2017)

126. Zaidan, O.F., Callison-Burch, C.: Arabic dialect identification. Computational Linguistics **40**(1), 171–202 (2014)

127. Zhang, L., Wang, S., Liu, B.: Deep learning for sentiment analysis: A survey p. e1253 (2018)

128. Zhang, X., Zhao, J., LeCun, Y.: Character-level convolutional networks for text classification. In: Advances in neural information processing systems, pp. 649–657 (2015)

129. Zhou, X., Wan, X., Xiao, J.: Attention-based lstm network for cross-lingual sentiment classification. In: Proceedings of the 2016 conference on empirical methods in natural language processing, pp. 247–256 (2016)

130. Zhou, Z.H., Feng, J.: Deep forest: Towards an alternative to deep neural networks. arXiv preprint arXiv:1702.08835 (2017)